CLAIMS: We claim:

1. A method for attracting the attention of people in public places and engaging them in an interaction for getting information and entertainment, comprising the states of;

a) a Wait State;

b) an Attraction State;

c) an User Engagement State;

d) an User Interaction State; and

e) an Interaction Termination State.

2. The method according to claim 1, wherein the states comprise a plurality of modules,

whereby the module is defined as a standalone application or a collection of applications,

whereby said collection of applications is a container of sub-applications, and

whereby said collection of applications can also manage the execution of individual sub-applications.

3. The method according to claim 1, wherein the selection of said information and entertainment content is achieved by a touch-free interaction.

4. The method according to claim 3, wherein said touch-free interaction modality is fused with speech recognition input modality to provide multi-modality.

5. The method according to claim 1, wherein said information and entertainment content in a hierarchical structure is delivered by a multimedia display system.

6. The method according to claim 5, wherein said information and entertainment content is updated using a network.

7. The method according to claim 1, wherein the user sensing and hand motion sensing is processed using one or a plurality of image-capturing systems and a set of computer vision algorithms.

8. The method according to claim 7, wherein said image-capturing system divides its field of view into a plurality of capture zones, and apply said set of computer vision algorithms within said capture zones to sense said user and its hand motion.

9. The method according to claim 8, wherein said user is promoted or demoted depending on the coordinate of said user's position in said plurality of capture zones of said image-capturing system.

10. The method according to claim 7, wherein said image-capturing system is able to control its functionalities, such as pan, tilt, zoom, focus, auto-exposure, and white balance, if said image-capturing system is equipped with these functionalities, according

to the applications of a specific embodiment in order to adapt to the environment and said user's uniqueness.

11. The method according to claim 1, wherein the essentials of the background technology based on the computer vision algorithms further comprises the "3-I Technologies", the Intelligence Technologies, the Interaction Technologies, and the Immersive Technologies.

12. The method according to claim 11, wherein said Intelligence Technologies comprises means for data collection of said user, means for gathering usage statistics, means for getting user demographics, means for processing real-time personalization based on said demographics, means for processing security applications for authentication based on face recognition.

13. The method according to claim 11, wherein said Interaction Technologies comprises means for face/body detection, means for localization, means for tracking, means for gesture recognition and means for multi-modal integration to facilitate interaction with digital content, and means for appliances for information access and entertainment.

14. The method according to claim 11, wherein said Immersive Technologies comprises passive immersion, active immersion, and mixed immersion (Augmented Reality),

whereby said *passive immersion* integrates said user's face and body into an application or video, while said user(s) face/body is immersed into selectable background, such as group photos in college campus background,
whereby said *active immersion* allows said users to control avatars constructed using said user's face image, and
whereby said *mixed immersion* (Augmented Reality) allows said users to virtually interact with virtual objects.

15. The method according to claim 1, wherein said Wait State further comprises:
means for playing video loop for advertising purpose and playing content intended to run in an introduction state.

16. The method according to claim 1, wherein said Attraction State further comprises:
means for attracting people and engaging them to the interaction with the embodiment, whereby the attraction means is an active and intelligent way of interrupting said user by graphical effects, sound effects, or mechanical effects, encouraging said user to engage in the interaction with said invention.

17. The method according to claim 1, wherein said User Engagement State further comprises:
means for helping said users to engage in an interaction with the embodiment smoothly by training them to know how to use said embodiment,

whereby the training methods can be in the graphical, the vocal, and the literal forms.

18. The method according to claim 1, wherein said User Interaction State further comprises:

means for providing said information and entertainment content to said user.

19. The method according to claim 1, wherein said User Interaction State further comprises:

means for providing the interaction not only to a single user but also to a plurality of users (crowd).

20. The method according to claim 1, wherein said Interaction Termination State further comprises:

means for collecting data about said user,

whereby the method uses explicit and implicit data collection about said user and stores the result in a database,

whereby said implicit data collection is done by the computer vision based technologies in the method automatically,

whereby said data collection method in said Interaction Termination State is an explicit data collection.

21. An apparatus for attracting the attention of people in public places and engaging them in an interaction for getting information and entertainment, comprising;

a) one or a plurality of image-capturing system, which consist of one or a plurality of image-capturing devices;

b) output display system;

c) processing and controlling system;

d) sound system;

e) microphone;

f) lighting system; and

g) user interaction area.

22. The apparatus according to claim 21, wherein said image-capturing system is able to control its functionalities, such as pan, tilt, zoom, focus, auto-exposure, and white balance, if said image-capturing system is equipped with these functionalities, according to the applications of the specific embodiment in order to adapt to the environment and said user's uniqueness.

23. The apparatus according to claim 21, wherein said processing and controlling system comprises one or multiple processors performing the steps of:

a) handling and processing a plurality of computer vision algorithms;

b) detecting and tracking face images from the continuous video input image sequences;

c) detecting and tracking the hand image from the continuous video input image sequences;

d) loading graphical images to the video memory space and displaying the content on the display system;

e) processing speech recognition; and

f) handling the interaction between said user and the system.

24. The apparatus according to claim 21, wherein said image-capturing system divides its field of view into plurality of capture zones, and apply said set of computer vision algorithms within said capture zones to sense said user and its hand motion.

25. The apparatus according to claim 21, wherein said user is promoted or demoted depending on the coordinate of said user's position in said plurality of capture zones of said image-capturing system.

26. The apparatus according to claim 23, wherein the step a) handling and processing a plurality of computer vision algorithms further comprises the "3-I Technologies", the Intelligence Technologies, the Interaction Technologies, and the Immersive Technologies.

27. The apparatus according to claim 26, wherein the apparatus further comprises means for handling said Intelligence Technologies, which comprise means for data

collection of said user, means for gathering usage statistics, means for getting user demographics, means for processing real-time personalization based on said demographics, means for processing security applications for authentication based on face recognition.

28. The apparatus according to claim 26, wherein the apparatus further comprises means for handling said Interaction Technologies, which comprise means for face/body detection, means for localization, means for tracking, means for gesture recognition and means for multi-modal integration to facilitate interaction with digital content, and means for appliances for information access and entertainment.

29. The apparatus according to claim 26, wherein the apparatus further comprises means for handling said Immersive Technologies, which comprise passive immersion, active immersion, and mixed immersion (Augmented Reality),
whereby said passive immersion integrates said user's face and body into an application or video, while said user(s) face/body is immersed into selectable background, such as group photos in college campus background,
whereby said active immersion allows said users to control avatars constructed using said user's face image, and
whereby said mixed immersion (Augmented Reality) allows said users to virtually interact with virtual objects.

30. An apparatus for attracting the attention of people in public places and engaging them in an interaction for getting information and entertainment, comprising the states of;

a) a Wait State;

b) an Attraction State;

c) an User Engagement State;

d) an User Interaction State; and

e) an Interaction Termination State.

31. The apparatus according to claim 30, wherein said Wait State further comprises: means for playing video loop for advertising purpose and playing content intended to run in an introduction state.

32. The apparatus according to claim 30, wherein said Attraction State further comprises:

means for attracting people and engaging them to the interaction with the embodiment, whereby the attraction means is an active and intelligent way of interrupting said user by graphical effects, sound effects, or mechanical effects, encouraging said user to engage in the interaction with said invention.

33. The apparatus according to claim 30, wherein said User Engagement State further comprises:

means for helping said users to engage in an interaction with the embodiment smoothly by training them to know how to use said embodiment,

whereby the training methods can be in the graphical, the vocal, and the literal forms.

34. The apparatus according to claim 30, wherein said User Interaction State further comprises:

means for providing said information and entertainment content to said user.

35. The apparatus according to claim 30, wherein said User Interaction State further comprises:

means for providing the interaction not only to a single user but also to a plurality of users (crowd).

36. The apparatus according to claim 30, wherein said Interaction Termination State further comprises:

means for collecting data about said user,

whereby the apparatus uses explicit and implicit data collection about said user and stores the result in a database,

whereby said implicit data collection is done by said computer vision based technologies in the apparatus automatically,

whereby said data collection method in said Interaction Termination State is an explicit data collection.